

Real-time Tooth Region Detection in Intraoral Scanner Images with Deep Learning

Na-Yun Park.* · Ji-Hoon Kim* · Tae-Min Kim* · Kyeong-Jin Song* · Yu-Jin Byun* · Min-Ju Kang* ·
Kyungkoo Jun** · Jae-Gon Kim*[†]

*Department of Industrial and Management Engineering, Incheon National University

**Department of Embedded Systems Engineering, Incheon National University

딥러닝을 이용한 구강 스캐너 이미지 내 치아 영역 실시간 검출

박나윤* · 김지훈* · 김태민* · 송경진* · 변유진* · 강민주* · 전경구** · 김재곤*[†]

*인천대학교 산업경영공학과, **인천대학교 임베디드시스템공학과

In the realm of dental prosthesis fabrication, obtaining accurate impressions has historically been a challenging and inefficient process, often hindered by hygiene concerns and patient discomfort. Addressing these limitations, Company D recently introduced a cutting-edge solution by harnessing the potential of intraoral scan images to create 3D dental models. However, the complexity of these scan images, encompassing not only teeth and gums but also the palate, tongue, and other structures, posed a new set of challenges. In response, we propose a sophisticated real-time image segmentation algorithm that selectively extracts pertinent data, specifically focusing on teeth and gums, from oral scan images obtained through Company D's oral scanner for 3D model generation. A key challenge we tackled was the detection of the intricate molar regions, common in dental imaging, which we effectively addressed through intelligent data augmentation for enhanced training. By placing significant emphasis on both accuracy and speed, critical factors for real-time intraoral scanning, our proposed algorithm demonstrated exceptional performance, boasting an impressive accuracy rate of 0.91 and an unrivaled FPS of 92.4. Compared to existing algorithms, our solution exhibited superior outcomes when integrated into Company D's oral scanner. This algorithm is scheduled for deployment and commercialization within Company D's intraoral scanner.

Keywords : Computer Vision, Deep Learning, Image Segmentation, Oral Scanner

1. 서론

의학 기술 발달의 영향으로 현대인의 평균 수명이 증가함에 따라 현대인의 삶의 질을 개선시키는 치아 건강은 중요한 요소로 여겨진다. 구강질환 및 외상에 의한 치아의 훼손은 저작 활동, 발음 등의 기능 수행에 어려움을 초래

한다[10]. 이에 치의학 분야에서는 치과 보철물을 제작하여 환자의 구강 기능을 회복시키고 구강 건강을 유지하게 한다. 전통적인 치과 기공 작업이 노동집약적이고 경험의존적이라는 문제를 해결하기 위해 1980년대 이후 치과분야에서 CAD/CAM(Computer-Aided Design/Computer-Aided Manufacturing) 기술이 도입되었다. 초기에는 치아를 구강 스캐너로 완전히 디지털화하는 기술이 발달되지 않았기에 인상 채득 후 석고 모형을 제작하여 스캔해 CAD/CAM을 활용하는 시스템이 보편적이었다[3]. 기존 인상 채득 과정은 트레이에 인상재를 넣어 구강 내 형상을 인상체로 제작

Received 1 September 2023; Finally Revised 13 September 2023;
Accepted 13 September 2023

[†] Corresponding Author : jaegkim@inu.ac.kr

하고 모형제를 주입하여 작업 모형을 제작하는 아날로그 인상 시스템이다. 이 과정은 인상재와 트레이가 필수적이며 인상재의 변형, 감염, 환자의 불편함 등을 유발하고 치기공사의 풍부한 임상적 경험을 요구한다.

이러한 문제점들을 해결하기 위해 구강 스캐너를 활용하여 인상을 채득하는 디지털 인상 채득 기술이 도입되었다 [12]. 치과용 CAD/CAM 시스템을 통한 디지털 보철물 제작은 세 단계로 진행된다. 첫 번째 단계는 구강 내의 치아를 3차원적으로 스캔하는 과정(Data Capture Process)이며 다음으로 스캐닝한 데이터를 바탕으로 보철물을 디자인하는 CAD 과정(Design Process), 마지막으로 CAD 소프트웨어에서 디자인한 보철물을 출력하는 CAM과정(Manufacturing Process)이 수행된다. 특히 보철물이 채득된 인상에 의거하여 제작되므로 구강 내를 스캔하는 과정은 보철물 제작에 있어 가장 기초적이며 중요한 단계이다[8, 9]. 구강 스캐너는 실시간으로 구강 내를 스캔하여 3D 치아 모형을 생성한다. 스캔하여 얻은 이미지로부터 3D 치아 모형을 생성하기 위해서는 치아와 잇몸을 제외한 혀, 입천장 등의 노이즈를 분리하여 제거하는 후처리 알고리즘이 필요하다.

국내 기업 D사는 구강 스캐너와 구강 밀링머신을 개발하는 덴탈 CAD/CAM 토탈 솔루션 기업이다. 2022년에 D사에서는 디지털 인상 채득이 가능한 구강 스캐너를 개발하였다. 이 때 후처리 알고리즘으로 구강 스캐너에서 촬영된 이미지와 깊이(Depth) 정보만을 활용하여 1차적으로 노이즈를 구분한 뒤 치아 영역만을 마킹하는 로직을 사용하고 있다. 하지만 이는 혀, 입천장 등의 노이즈를 정밀히 제거하지 못할 뿐 아니라 치아의 형태를 보존하지 못하는 실정이다. 이로 인해 수작업으로 제거되지 않은 노이즈를 처리하는 작업을 거치고 있다. 따라서 본 연구에서는 구강 이미지에서 노이즈를 제거하는 딥러닝 기반 이미지 분할(Image Segmentation) 알고리즘을 개발하고자 한다.

이미지 분할은 원본의 이미지에서 특정 대상을 분할하는 기술이다. 개인마다 치아의 모양과 배열이 다르기 때문에 이를 치과 분야에 적용하기는 쉽지 않다[11, 18]. 하지만 알파고의 등장과 함께 다가온 4차 산업혁명의 영향으로 딥러닝 기술이 급격히 발전했으며 이는 이미지 분할에 적용되어 두각을 드러내고 있다[14]. Bae et al.[1]은 치과 분야에서도 인공지능을 적절히 활용했을 때 치아의 분류 및 분석이 가능하다고 제안하였으며 Corbella et al.[4]은 다양한 치과 분야의 28개의 연구에서 딥러닝을 적용했을 때 기존의 알고리즘에 비해 높은 정확도를 보인다고 제시하였다. 이처럼 최근 치과 분야의 딥러닝 기반 이미지 분할 기법 적용은 많은 연구자들에게 주목받고 있다.

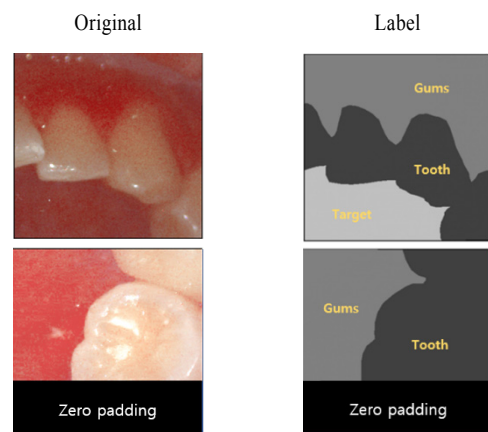
주로 치과 분야의 이미지 분할 알고리즘은 X-ray 또는 CT 이미지를 활용한다[13, 16]. 하지만 본 연구에서는 실제 구강 스캔 이미지를 연구에 사용하였다. 실제 구강 스

캔 이미지를 사용한 연구를 제안한 Zhu et al.[19]은 mask R-CNN 모델을 활용한 치아 이미지 분할 알고리즘을 개발하여 픽셀 정확도 약 95%를 달성하였다. 하지만 해당 모델은 평균적으로 약 5 FPS의 속도를 가진다. 실시간 구강 스캔이 이루어지는 환경에서는 모델의 정확도 뿐만 아니라 속도도 매우 중요한 평가 요소이다. 개발된 알고리즘은 정확도와 더불어 실시간 구강 스캔에 영향을 주지 않는 빠른 속도를 제공해야 한다. 따라서 실시간으로 이미지를 처리하는 구강 스캐너에 이를 적용하기는 어렵다. 한편 Xu et al.[15]에 의해 제안된 실시간 이미지 분할 알고리즘인 PIDNet은 속도와 정확도의 상충관계에서 두 요소를 모두 최대화한다. 특히 PIDNet-S는 100 FPS 이상의 속도를 제공하며 실시간으로 이미지를 분할할 수 있음이 제시되었다. 따라서 정확도와 속도 두 가지 요소의 절충점을 찾기 위해 PIDNet을 본 연구에 활용할 모델로 선정하였다.

2. 데이터셋 및 방법론

2.1 데이터셋

본 연구에서는 D사에서 제공한 약 2,400장의 구강 스캔 이미지를 bmp 형식으로 수집하였고 이를 8:2의 비율로 분할하여 학습 데이터로 1976장, 검증 데이터로 495장을 사용하였다. 학습을 위해 Intel에서 개발한 오픈소스 툴인 CVAT을 사용하여 이미지의 각 픽셀에 카테고리를 부여하는 라벨링 작업을 진행하였다. 치아를 의미하는 라벨을 Tooth, 잇몸을 Gums로 지정하고 그 외 혀, 입천장 등 제거가 필요한 영역을 Target으로 지정하였다. 또한 수집한 이미지 데이터의 크기가 일정하지 않은 문제를 해결하기 위해 좌상단을 기준으로 Zero padding을 적용하는 작업을 거쳐 <Figure 1>과 같이 행과 열이 모두 384개의 픽셀로 구성되도록 전처리하였다.



<Figure 1> Examples of Data Labeling

2.2 이미지 분할 모델

실시간으로 구강 스캔 이미지를 처리해야하는 구강 스캐너 특성상 높은 정확도와 더불어 빠른 처리 속도가 요구된다 [6]. 모델에 이미지를 입력하면 픽셀별로 클래스 레이블이 출력되며 동작 방식은 <Figure 2>와 같다. PIDNet controller는 세 개의 요소를 포함한다. Proportional(P) controller는 현재의 오류를 나타내며 Integral(I) controller는 과거의 오류를 수집한다. Derivate(D) controller는 미래의 오류 변화를 예측한다. 따라서 P, I, D는 각각 detailed, context, boundary branch를 의미한다. 이는 서로 상호보완적인 역할을 하고 기존의 2 branch network를 기반으로 하는 이미지 분할 알고리즘에서 주로 나타나던 overshoot 문제를 합리적으로 해결한다. 이러한 각각의 P, I, D branch를 손실 함수에 적용하여 네트워크를 학습시킨다. 이를 통해 입력과 출력 사이의 오차를 줄여 높은 성능을 이끌어낸다.

Zhao et al.[17]은 feature 추출 시 PPM(Pyramid Pooling Module)을 도입하여 기존 이미지 분할 알고리즘의 픽셀 분류 정확도를 향상시켰다. PPM은 Local 정보 뿐 아니라 Global context 정보를 활용하도록 하는 네트워크이다. Hong et al.[7]은 PPM의 구조를 개선하여 더 나은 성능을 보이는 DAPPM(Deep Aggregation Pyramid Pooling Module)을 제안하였다. 이를 통해 이미지 분할을 위하여 픽셀을 분류할 때 전체적인 문맥을 고려할 수 있고 인간이 통상적으로 가지고 있는 개념을 반영할 수 있다.

PIDNet은 학습 과정을 경량화하여 속도를 향상시키기 위해 DAPPM에서 채널의 수를 줄인 PAPP (Parallel Aggregation PPM)을 도입하였다. PAPP은 PIDNet-S, PIDNet-M에 적용되어 각각의 채널 수를 128에서 96으로 줄여 모델의 속도 향상에 기여하였다. PIDNet-L은 여전히 DAPPM을 사용하지만 채널의 수를 128에서 112로 줄이도록 설계되었다. PIDNet-S, PIDNet-M, PIDNet-L의 총 파라미터의 수는 각각 7.6M, 34.4M, 36.9M개로 설계되었다. 파라미터의 수가 적을수록 모델의 속도는 빨라지며 정확도는 감소하는 경향이 있으며 파라미터의 수가 많을수록 모델의 속도는 느려지지만 정확도는 향상되는 경향이 있다. 즉 PIDNet-S의 속도가 가장 빠르며 PIDNet-L의 속도가 가장 느리다. 본 연구에서는 PIDNet-S, PIDNet-M, PIDNet-L 모델을 활용하여 구강 스캔 이미지에서 실시간으로 노이즈를 제거하는 알고리즘을 개발하였다. 모델 학습을 위한 컴퓨터의 그래픽카드는 nvidia geforce rtx 3090, CPU는 Intel(R) Core(TM) i9-10900xCPU@ 3.70GHz를 사용하였고 GPU RAM 크기는 192GB이다. 사전 테스트를 통과해 최종 설정한 학습 파라미터는 <Table 1>과 같다.

<Table 1> Parameters for training PIDNet

Parameter	Value
Batch size	6
Epoch	300
Learning rate	0.001
Weight decay	0.0005

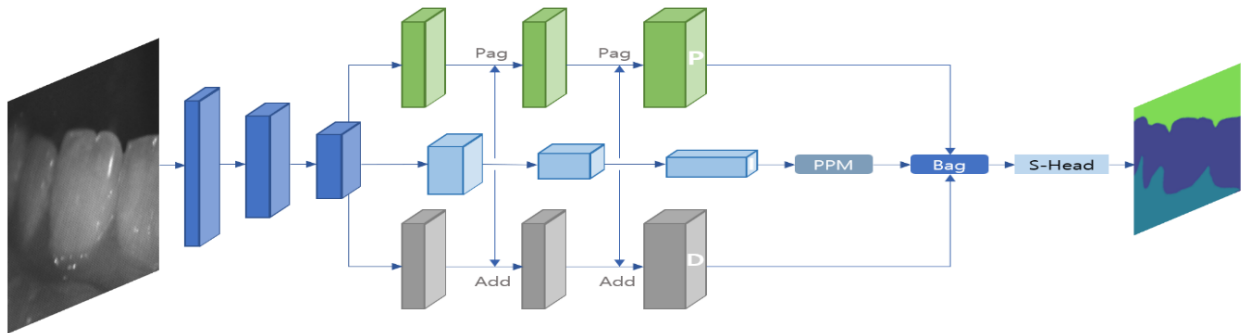
3. 실험 결과

실험 결과는 <Table 2>와 같다. 평가지표는 이미지 분야에서 주로 사용되는 정확도의 척도인 mIOU(Mean Intersection over Union)를 활용하였고 속도는 초당 프레임 수를 의미하는 FPS로 나타내었다[2]. D사에서 구강 스캐너에 해당 알고리즘을 탑재하기 위해 요구한 성능은 mIOU 85% 이상, FPS 80 이상이다. 이를 만족시키기 위해 하이퍼 파라미터를 조정하여 실험한 결과 PIDNet-S의 mIOU가 89.3%, FPS가 90.3로 모든 성능 지표에서 가장 우수한 성능을 보였다. 모델의 크기가 커짐에 따라 FPS가 점차 감소하였지만 mIOU도 동시에 감소하였다. PIDNet-M, PIDNet-L의 파라미터 수는 PIDNet-S의 파라미터 수보다 많으므로 학습에 포크를 증강시켜야 좋은 성능을 낼 가능성이 높다. 따라서 해당 모델들의 학습 에포크를 늘려서 학습시킨다면 PIDNet-S보다 더 좋은 mIOU를 보일 수 있을 것으로 예상된다. 하지만 현재 학습된 모델보다 학습을 더 오래 시킨다면 D사에서 요구한 80FPS를 만족시키기 어렵다. 이에 D사에서는 현재 도출된 mIOU 89.3%는 구강 스캐너 탑재 요건인 mIOU 85% 이상을 충족하므로 더 이상의 추가 학습 없이 90.3 FPS의 속도를 내는 PIDNet-S를 구강 스캐너에 탑재하기를 희망하였다. 따라서 본 연구에서는 디지털 구강 스캔의 자동화를 구축하기 위하여 PIDNet-S를 D사의 구강스캐너에 탑재하고자 한다.

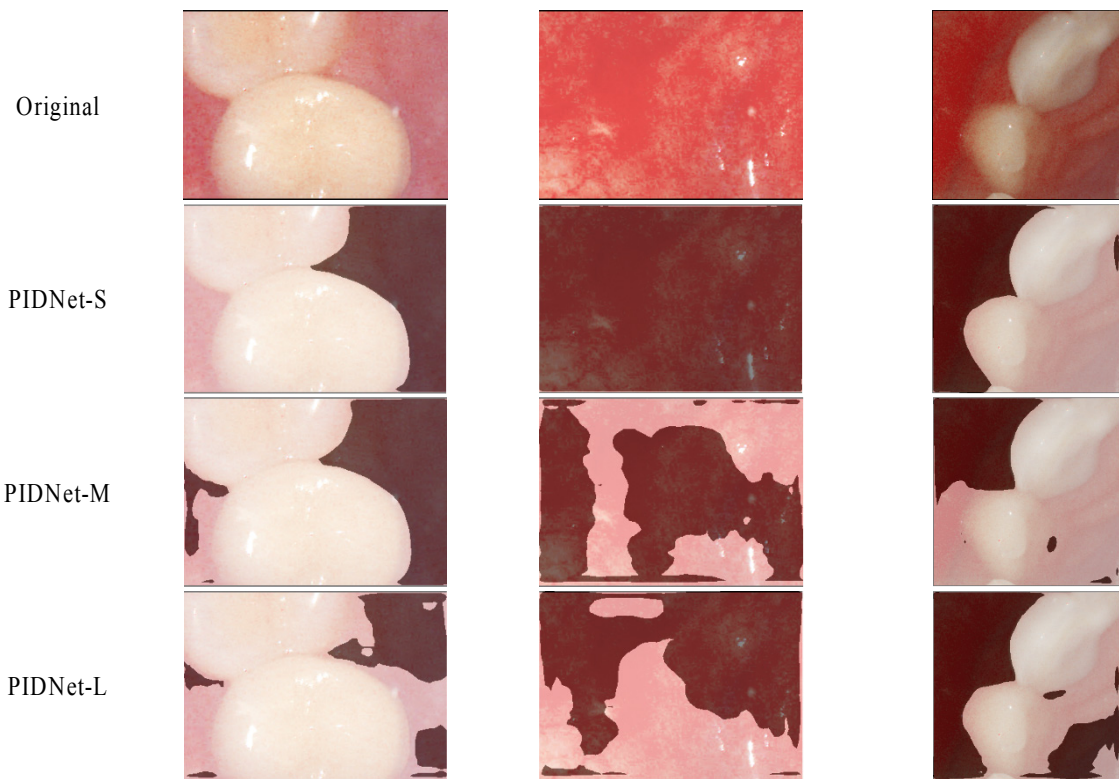
Original 이미지가 PIDNet-S, PIDNet-M, PIDNet-L 모델에 입력될 때의 각각의 이미지 분할 결과는 <Figure 3>과 같다. 어둡게 표시된 픽셀은 노이즈로 분할된 픽셀이며 밝게 표시된 픽셀은 치아 또는 잇몸으로 분할된 픽셀이다. 또한 <Figure 4>에서는 기존 알고리즘을 사용하여 치아 3D 모델을 정합한 결과에 비해 본 연구에서 개발한 알고리즘을 탑재하여 정합한 결과가 좋은 성능을 보임을 알 수 있다.

<Table 2> Performance comparison of PIDNet-S, PIDNet-M and PIDNet-L

Model	mIOU (%)	FPS
PIDNet-S	89.3	90.3
PIDNet-M	88.3	88.1
PIDNet-L	87.7	81.0

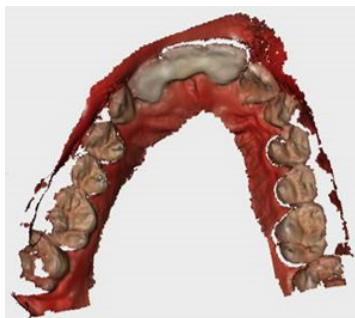


<Figure 2> Basic Architecture of PIDNet

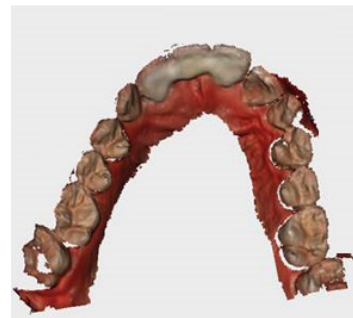


<Figure 3> Image Segmentation Results of PIDNet-S, PIDNet-M, and PIDNet-L Model

A result of the existing algorithm



A result of the suggested PIDNet-S



<Figure 4> Comparison of the results of the existing algorithm and PIDNet-S

4. 결론

본 연구에서는 D사의 구강 스캐너로부터 얻은 구강 스캔 이미지를 활용하여 3D 치아 모형을 생성하기 위한 이미지 분할 알고리즘을 개발하였다. 기존에는 치과 보철물 제작을 위해 인상 채득이 필요했으나, 이는 위생적인 문제와 환자의 불편함 등 여러 가지 측면에서 비효율적이다. 이에 본 연구에서 제안한 알고리즘은 실시간으로 노이즈를 제거하여 구강 스캔 이미지에서 유의미한 데이터인 치아와 잇몸을 추출한다. 딥러닝 기반의 실시간 이미지 분할 기술을 활용하여 해당 알고리즘을 개발하였다. 구강 스캔 이미지에는 치아와 잇몸뿐만 아니라 입천장, 혀 등 다른 요소들도 포함되어 있기 때문에, 개발된 알고리즘은 해당 부분을 제외하고 치아와 잇몸만을 추출한다. 이를 위해 약 2,400장의 구강 스캔 이미지를 활용하여 PIDNet-S, PIDNet-M, PIDNet-L 세 가지 모델을 각각 학습시켰고 그 중 D사에서 요구한 정확도와 속도의 성능을 가장 효율적으로 만족시킬 수 있는 PIDNet-S를 D사의 구강 스캐너 탑재를 위해 선정하였다.

본 연구에서 개발한 실시간 이미지 분할 모델은 mIOU 89.3%, FPS 90.3으로 정확도와 속도 측면에서 모두 높은 성능을 보여주며, 실시간으로 구강을 스캔하여 치아 3D 모델을 재구성함에 있어 노이즈 없이 원만한 정합이 가능하도록 해 줄 수 있다. 제안된 알고리즘은 D사의 구강 스캐너에 탑재되어 상용화될 예정이다. 현재에는 모델 학습을 위해 약 2,400장의 이미지 데이터만을 활용하였다. 추후에 구강 스캔 이미지 데이터를 추가로 수집하고 이를 재학습 시킴으로써 실시간 이미지 분할 성능을 계속해서 향상시킬 수 있을 것이다.

Acknowledgement

This work was supported by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (P0008691, HRD Program for Industrial Innovation).

References

- [1] Bae, E., Jeong, J., Son, Y., and Lim, J., A study on virtual tooth image generation using deep learning-based on the number of learning, *Journal of Technologic Dentistry*, 2020, Vol. 42, No. 1, pp. 1-8.
- [2] Behera, S.K., Rath, A.K., and Sethy, P.K., Fruits yield estimation using Faster R-CNN with MIoU, *Multimedia Tools and Applications*, 2021, Vol. 80, pp. 19043-19056.
- [3] Choi, H.S., Moon, J.E., and Kim, S.H., The application of CAD/CAM in dentistry, *The Journal of the Korean Dental Association*, 2012, Vol. 50, No.3, pp. 110-117.
- [4] Corbella, S., Srinivas, S., and Cabitza, F., Applications of deep learning in dentistry. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 2021, Vol. 132, No. 2, pp. 225-238.
- [5] He, K., Gkioxari, G., Dollár, P., and Girshick, R., Mask r-cnn, In *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961-2969.
- [6] Holder, C.J. and Shafique, M., On efficient real-time semantic segmentation: A survey, <https://doi.org/10.48550/arXiv.2206.08605>.
- [7] Hong, Y., Pan, H., Sun, W., and Jia, Y., Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes, arXiv preprint arXiv:2101.06085.
- [8] Huang, M. Y., Son, K., Lee, W. S., and Lee, K. B., Comparison of the accuracy of intraoral scanner by three-dimensional analysis in single and 3-unit bridge abutment model: In vitro study, *The Journal of Korean Academy of Prosthodontics*, 2019, Vol. 57, No. 2, pp. 102-109.
- [9] Kim, J.H., Kim, J.H., and Kim, H.Y., A study on common errors in digital impressions:(An Example of CEREC $\text{\textcircled{R}}$ $\text{\textcircled{S}}$ AC), *Journal of Technologic Dentistry*, 2011, Vol. 33, No. 3, pp. 211-218.
- [10] Kim, Y.G., A study on the life expectancy of the dental prosthesis, *Journal of Oral Medicine and Pain*, 1995, Vol. 20, No. 2, pp. 317-325.
- [11] Liu, X., Song, L., Liu, S., and Zhang, Y., A review of deep-learning-based medical image segmentation methods, *Sustainability*, 2021, Vol. 13, No. 3, pp. 1224.
- [12] Pyo, S.W., Park, Y.B., Kim, J.H., Moon, H.S., and Lee, K.W., Maxillary anterior all ceramic restoration using digital impression and CAD/CAM, *The Journal of Korean Academy of Prosthodontics*, 2011, Vol. 49, No. 3, pp. 263-269.
- [13] Rad, A.E., Rahim, M.S., and Norouzi, A., Digital dental X-ray image segmentation and feature extraction, *Telkomnika Indonesian Journal of Electrical Engineering*, 2013, Vol. 11, No. 6, pp. 3109-3114.
- [14] Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van, D., and Hassabis, D., Mastering the game of Go with deep neural networks and tree search, *Nature*,

2016, Vol. 529, No. 7587, pp. 484-489.

- [15] Xu, J., Xiong, Z., and Bhattacharyya, S.P., PIDNet: A real-time semantic segmentation network inspired by PID controllers, In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 19529-19539.
- [16] Yang, Y., Xie, R., Jia, W., Chen, Z., Yang, Y., Xie, L., and Jiang, B., Accurate and automatic tooth image segmentation model with deep convolutional neural networks and level set method, *Neurocomputing*, 2021, Vol. 419, pp. 108-125.
- [17] Zhao, H., Shi, J., Qi, X., Wang, X. and Jia, J., Pyramid scene parsing network, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2881-2890.
- [18] Zhao, M., Ma, L., Tan, W., and Nie, D., Interactive tooth segmentation of dental models, In *2005 IEEE*

Engineering in Medicine and Biology 27th Annual Conference, 2006, pp. 654-657.

- [19] Zhu, G., Piao, Z., and Kim, S.C., Tooth detection and segmentation with mask R-CNN, In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, 2020, pp. 070-072.

ORCID

- Na-Yun Park | <https://orcid.org/0009-0003-1532-0285>
 Ji-Hoon Kim | <https://orcid.org/0009-0009-6548-3721>
 Tae-Min Kim | <https://orcid.org/0000-0003-2400-3161>
 Kyeong-Jin Song | <https://orcid.org/0009-0004-1945-5627>
 Yu-Jin Byun | <https://orcid.org/0009-0000-1090-4482>
 Min-Ju Kang | <https://orcid.org/0009-0008-7983-2294>
 Kyungkoo Jun | <https://orcid.org/0000-0002-0080-4573>
 Jae-Gon Kim | <http://orcid.org/0000-0002-4821-4441>